Generic CBTS: Correlation based Transformation Strategy for Privacy Preserving Data Mining

Abstract

Mining useful knowledge from corpus of data has become an important application in many fields. Data Mining algorithms like Clustering, Classification work on this data and provide crisp information for analysis. As these data are available through various channels into public domain, privacy for the owners of the data is increasing need. Though privacy can be provided by hiding sensitive data, it will affect the Data Mining algorithms in knowledge extraction, so an effective mechanism is required to provide privacy to the data and at the same time without affecting the Data Mining results. Privacy concern is a primary hindrance for quality data analysis. Data mining algorithms on the contrary focus on the mathematical nature than on the private nature of the information. Therefore instead of removing or encrypting sensitive data, we propose transformation strategies that retain the statistical, semantic and heuristic nature of the data while masking the sensitive information. The proposed Correlation Based Transformation Strategy (CBTS) combines Correlation Analysis in tandem with data transformation techniques such as Singular Value Decomposition (SVD), Principal Component Analysis (PCA) and Non
Negative Matrix Factorization (NNMF) provides the intended level of privacy preservation and enables data analysis. The proposed technique will work for numerical, ordinal and nominal data. The outcome of CBTS is evaluated on standard datasets against popular data mining techniques with significant success and Information Entropy is also accounted.

References

11. Dowon Hong and Abedelaziz Mohaisen "Augmented Rotation-Based Transformation for Privacy-Preserving Data Clustering" ETRI Journal, Volume 32, Number 3, June 2010.


Index Terms

Computer Science                  Information Sciences

Keywords

Transformation Strategy, Privacy Preserving Data Mining, Correlation Analysis, Information Entropy